SUPPLEMENTARY MATERIAL

Methods - Metabolomics and bioinformatics

We used the metabolomics analysis pipeline outlined by Ahmed and colleagues (1). GCMS data were processed using Automated Mass Spectral Deconvolution and Identification System software (AMDIS, version 2.71, 2012), the National Institute of Standards and Technology (NIST) mass spectral library (version 2.0, 2011) and the R package Metab (R version 3.3.2, 2016; Metab version 1.8.0). AMDIS and NIST software were used to build a local VOC library. A forward and reverse match of 900/1000 and an area threshold greater than 1,000,000 were used for assigning tentative compound identifications. A set match criterion of >90% was used. Compounds were named using IUPAC nomenclature.

Metaboanalyst v. 5.0 was used for metabolomics statistical analysis (2). Compounds present in laboratory air samples, and those present in <50% of all groups were removed from the comparisons. AMDIS reports assign NA when a VOC was not detected in a sample; this was changed to 1, before logarithmic transformation, median normalisation and standardisation.

Methods - DNA extraction

Two separate DNA extraction kits were utilised for extraction of fungal microorganisms from participant samples - PSP Spin Stool DNA Kit (Stratec) and QIAamp Fast DNA Stool Mini Kit (Qiagen) (3, 4). The PSP and QIAamp kits have different sensitivities for fungi detection. We used two kits for extraction and combined the extracts to improve the overall sensitivity of our results. Extractions were performed using standard protocols, with an additional bead-beating step at the start of the PSP extraction (3, 4).

Extracted DNA was quantified by Qubit (Qubit dsDNA HS assay Kit, Life Technologies): 7 PD samples with low yield were discarded. A universal tail tag dual index barcoding approach was used for amplicon preparation, specifically an optimised 18S rRNA protocol described by Frau *et al (5)*. The forward primer was FungiQuant-F (GGR AAA CTC ACC AGG TCCA G) and reverse FungiQuant-R (GSW CTA TCC CCA KCA CGA) (6). Primer overhang and barcoded index primers are listed in supplementary table 4. The resulting amplicons underwent sequencing in the Centre for Genomic Research (CGR) at the University of Liverpool. After quantifying the pool of amplicons using Qubit and assessing fragment distribution using an Agilent DNA high sensitivity kit (2100 Bioanalyzer, Agilent), the pooled

library of amplicons was sequenced using an Illumina MiSeq platform. Control samples (n=20) were prepared and sequenced in the same way but in a separate batch.

Methods - Mycobiome sequencing and bioinformatics

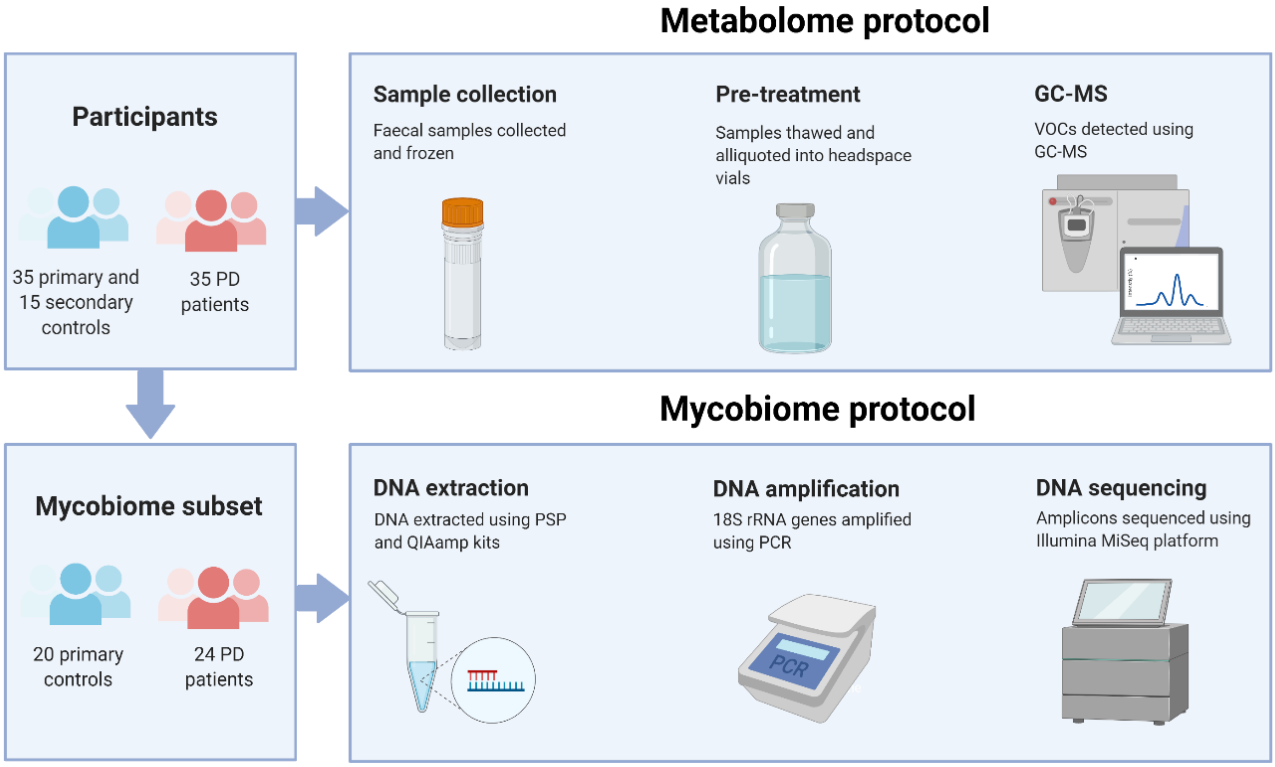
Mycobiome sequencing and analysis methodology used in this study was published in 2019 and the application of this method was published this year in gut microbes (5, 7). Sequencing data from the 18S rRNA amplicon pool was processed using the QIIME pipeline (v 1.9.1) (8). Sequenced reads underwent demultiplexing and adaptor and quality trimming (Cutadapt v 1.2.1 (9) and Sickle v 1.2 (10)– these steps were performed by the University of Liverpool Center for Genomic Research (CGR). Chimeric sequences were identified and filtered using UCHIME (11). Reads were error-corrected using the BayesHammer (12) module in SPAdes (v 3.7.9) (13) and pair-end read merging was carried out with PEAR (v 0.9.10) (14) and Phix reads were removed with a custom script. Reads were clustered with SWARM 2.0 (d = 3) (15). Using BLAST (16), taxonomy was assigned with assign_taxonomy.py in QIIME – we used SILVA (SILVA_123) (17) as our reference database. From this, an OTU table was generated using make_otu_table.py on QIIME. OTUs were filtered at a 0.05% threshold. Reads were aligned using the PyNAST (18) algorithm on QIIME and gaps (80% gap filter threshold) and hypervariable regions (10% most variable regions) removed using filter_allignment.py on QIIME. Finally, we created an approximately-maximum-likelihood phylogenetic tree using the FastTree2 (19) algorithm on QIIME. Reads from 15 OTUs were discarded due to non-fungal origin using filter_otus_from_otu_table.py on QIIME.

Methods references

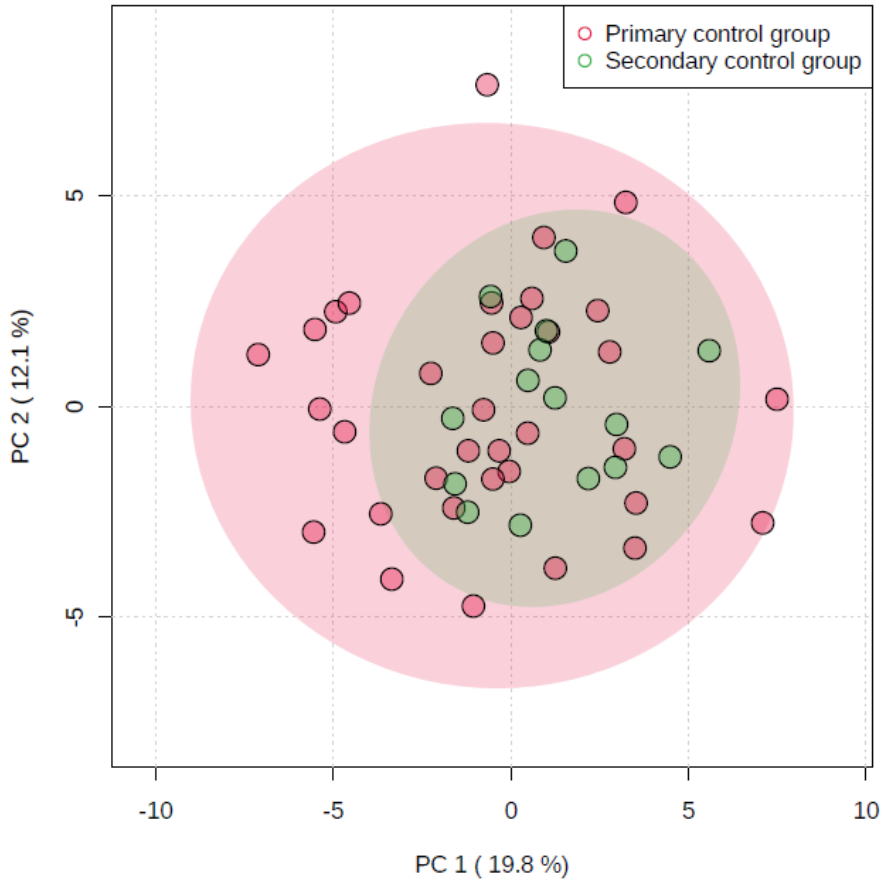
1. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Aliment Pharmacol Ther. 2016;43(5):596-611.
2. Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res. 2009;37(Web Server issue):W652-60.
3. "STRATEC Molecular GmbH". User manual PSP Spin Stool DNA Kit / PSP Spin Stool DNA Plus Kit 2013 [Available from: http://www.negev-bioproducts.com/Content/Images/CatalogImages/ProductVariations/PV_ID_51/PSPSpinStool_StoolPlusKit_23.pdf.
4. "Qiagen". QIAamp® Fast DNA Stool Mini Handbook 2020 [Available from: https://www.qiagen.com/gb/resources/resourcedetail?id=2a3f2c0b-2e8a-49fd-b442-829108ae1a4a&lang=en.
5. Frau A, Kenny JG, Lenzi L, Campbell BJ, Ijaz UZ, Duckworth CA, et al. DNA extraction and amplicon production strategies deeply influence the outcome of gut mycobiome studies. Sci Rep. 2019;9(1):9328.
6. Liu CM, Kachur S, Dwan MG, Abraham AG, Aziz M, Hsueh PR, et al. FungiQuant: a broad-coverage fungal quantitative real-time PCR assay. BMC Microbiol. 2012;12:255.

7. Frau A, Ijaz UZ, Slater R, Jonkers D, Penders J, Campbell BJ, et al. Inter-kingdom relationships in Crohn's disease explored using a multi-omics approach. Gut Microbes. 2021;13(1):1930871.
8. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.
9. Martin M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnetjournal. 2011;17(1):10-2.
10. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. 2011 [Available from: https://github.com/najoshi/sickle.
11. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194-200.
12. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC Genomics. 2013;14 Suppl 1(Suppl 1):S7.
13. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol. 2013;20(10):714-37.
14. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30(5):614-20.
15. Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ. 2015;3:e1420.
16. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460-1.
17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
18. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26(2):266-7.
19. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.
20. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, et al. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics. 2016;17(1):55.
21. Liu CM, Kachur S, Dwan MG, Abraham AG, Aziz M, Hsueh P-R, et al. FungiQuant: A broad-coverage fungal quantitative real-time PCR assay. BMC Microbiology. 2012;12(1):255.

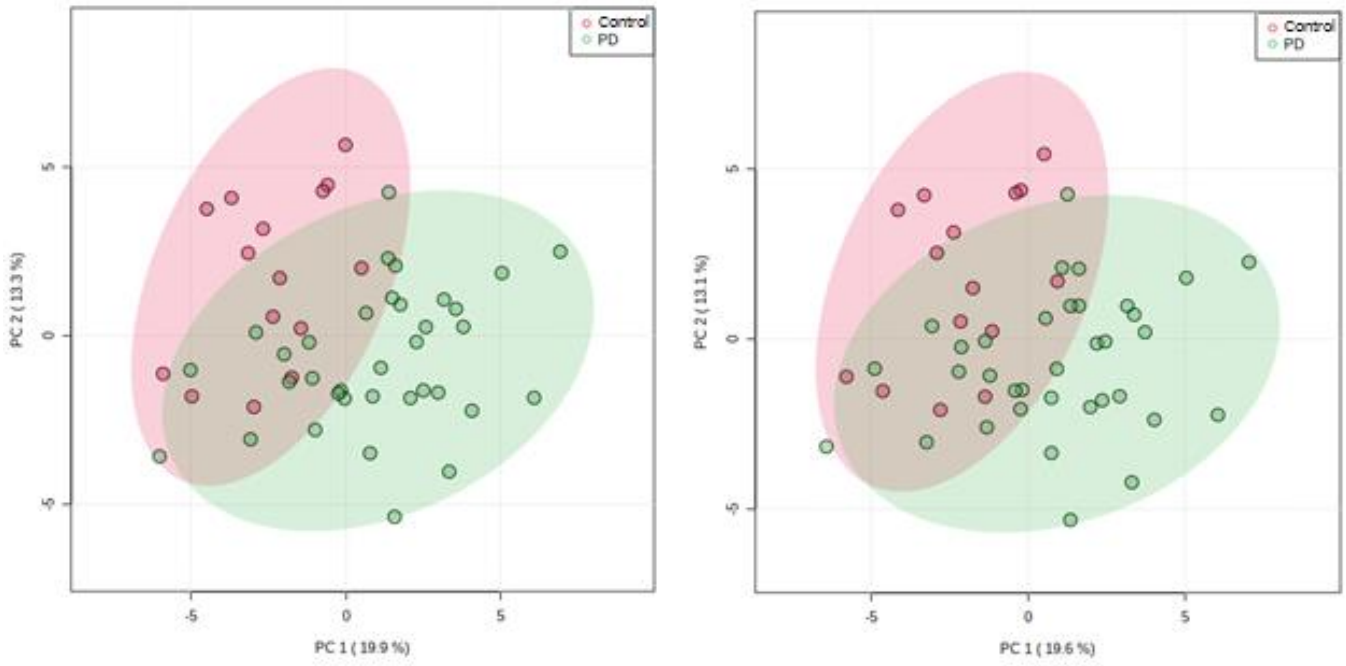
Supplementary figure 1. Summary of metabolome and mycobiome analyses (created with BioRender.com).



Supplementary figure 2. Principal component analysis (PCA) comparing the volatile organic compound profiles of primary and secondary control groups. Principal components analysis (PCA) is an unsupervised method that changes a range of potentially correlated variables (a similarity matrix) into a smaller range of uncorrelated variables called principal components – in doing so, it reduces the dimensions of complex data to enable visualization.

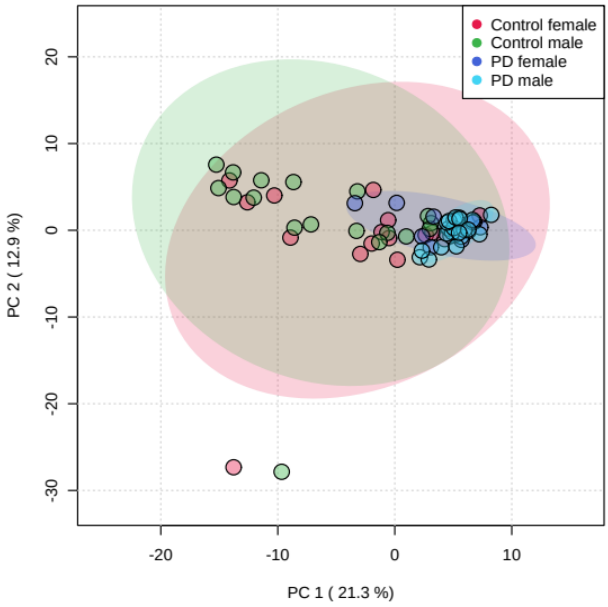


Supplementary figure 3. Principal component analysis (PCA) visualizing volatile organic compound profiles from PD patients and the secondary control group. There was no extra separation between PD and the second control group before (left) and after (right) removal of 1,3-ditert-butylbenzene from the PCA scatterplots.

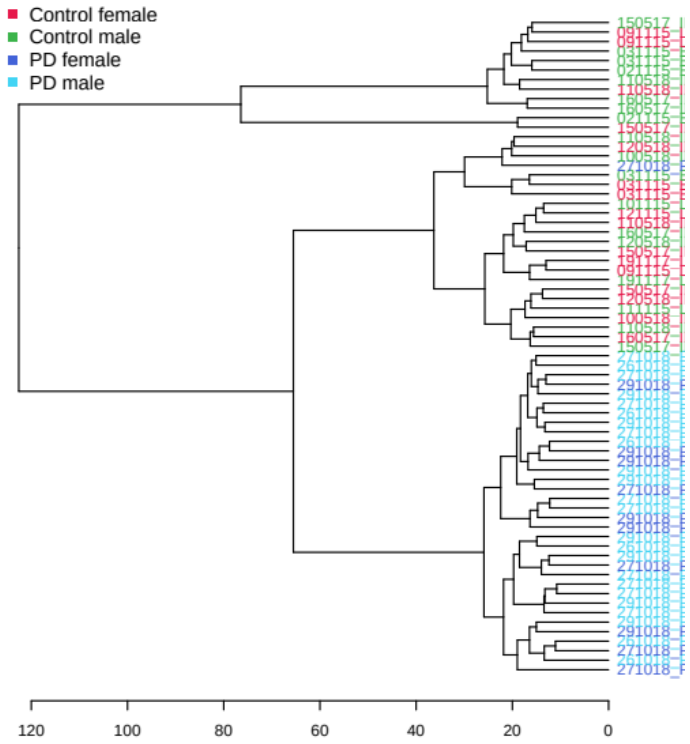


Supplementary figure 4. There was no significant difference between the VOC profiles of males and females in the PD and healthy control groups. **(A)** Two-dimensional PCA comparing males and females from the PD and control groups. **(B)** Dendrogram visualizing clustering within test groups.

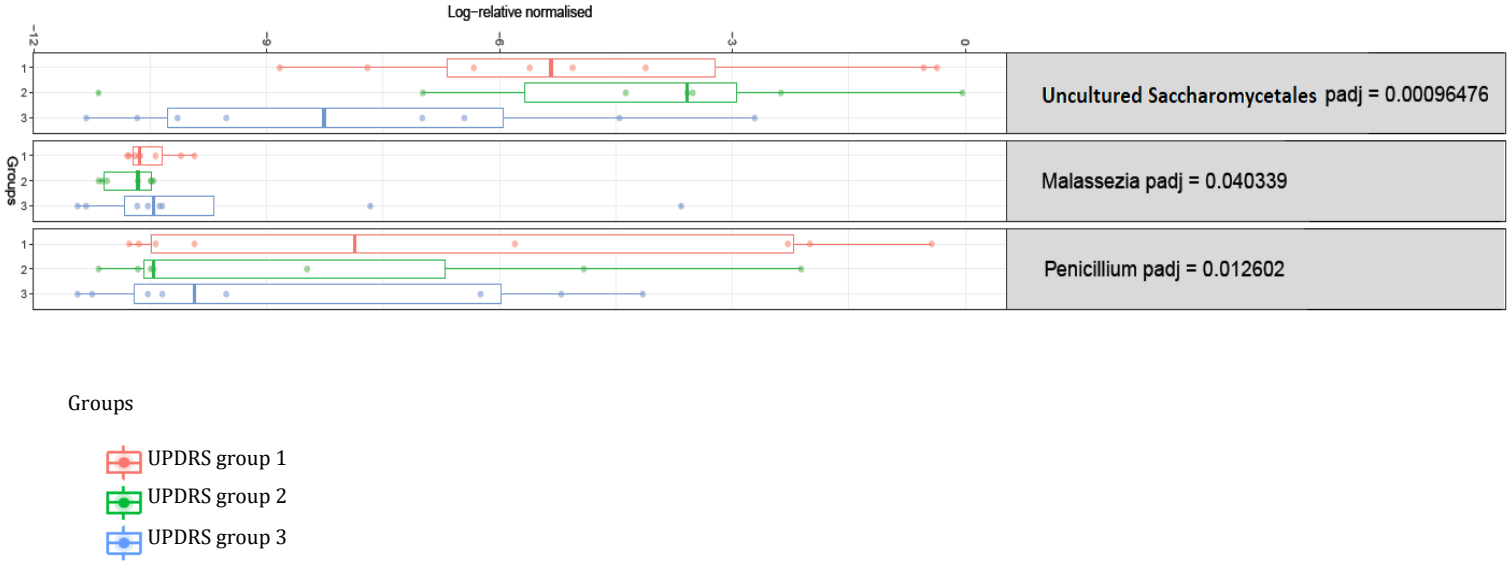
A.



B.

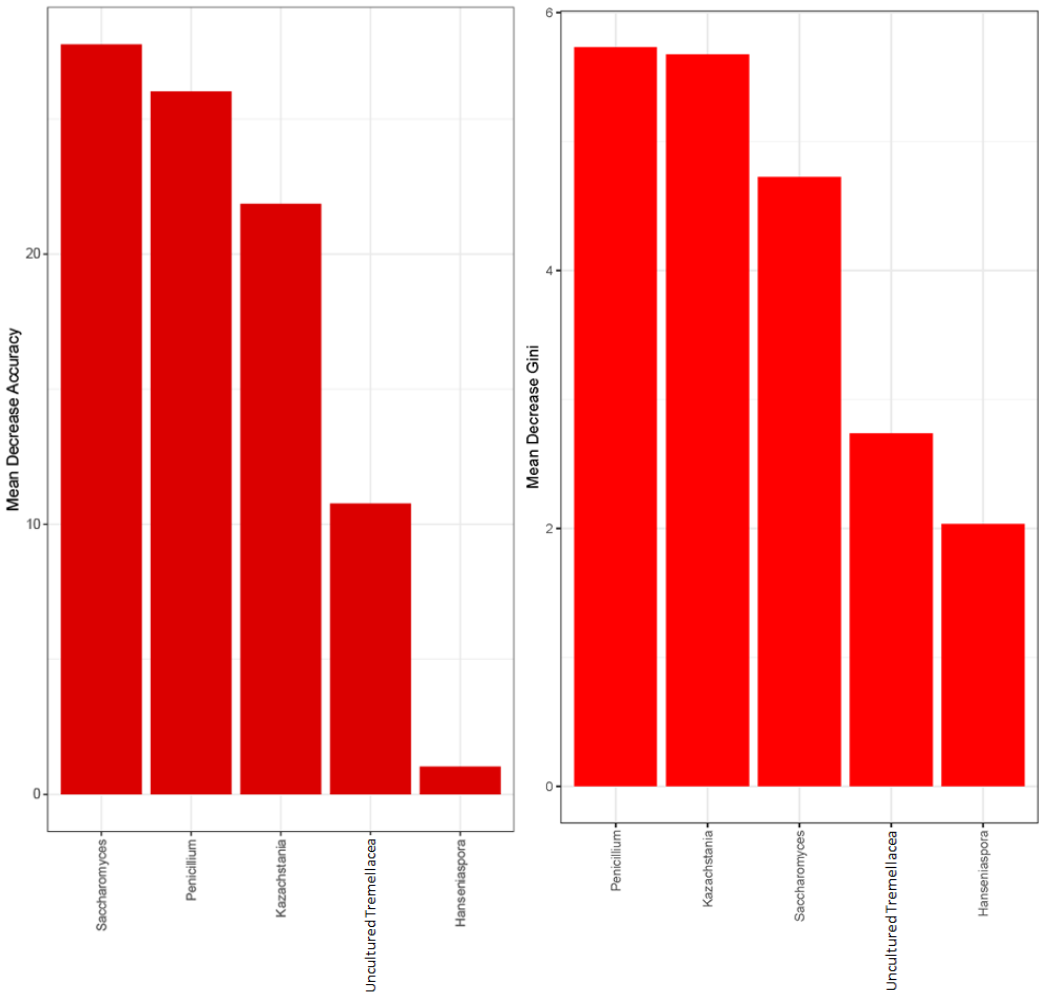


Supplementary figure 5. Box plot comparing the normalized abundances of genera significantly associated with PD severity (with abundance increasing as we go from left to right).



Supplementary material 6. Taxa differential analyses statistics and bar charts showing random forests classifier results. Mean decrease accuracy (MDA) indicates loss in accuracy of the model if that particular element is omitted. Mean decrease gini (MDG) indicates loss in purity of the model if that particular element is omitted.

	baseMean	log2FoldChange	pvalue	padj	Upregulated
Penicillium	427.6924804	7.593725577	5.04E-18	9.07E-17	PD
Saccharomyces	132626.6855	-2.341410535	1.16E-09	1.05E-08	Control
Kazachstania	5.520094777	2.709007225	3.29E-05	0.000197194	PD
Hanseniaspora	12.89941419	2.557021905	7.87E-05	0.000354238	PD
Uncultured Tremellaceae	873.1024652	3.153077821	0.000173312	0.000623924	PD



Supplementary Table 1. Summary of demographic and clinical features of participants.

Feature	Parkinson's disease	Primary control group	Secondary control group
Source	PD specialist clinic	Colonoscopy list	Healthy volunteers in diet study
Total number	35	35	15
Age: Range (Mean)	48-82 (67)	21-79 (61)	23-65 (48.2)
Gender	M:F = 24:11	M:F = 18:16	M:F = 10:5
Medication	Levodopa = 31; Dopamine agonist = 15; COMT inhibitor = 14; MAO-B inhibitor = 16; Amantadine = 17; Anticholinergic = 6; Laxatives = 17	Not known	Not known
MDS-UPDRS*: Range (Mean)	4-93 (27.4)	N/A	N/A
PACSYM†: Range (Mean)	2-36 (12.6)	N/A	N/A

*Part III of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (Goetz CG et al; Movement Disorders 2008).

† Patient Assessment of Constipation-Symptoms (PAC-SYM) – 12 item self-reported assessment of constipation symptoms (Frank L et al; Scand J Gastroenterol 1999).

Supplementary table 2. Summary of VOCs that had different abundance in primary and secondary control groups.

Volatile organic compound	FDR*	Increased in
Primary controls vs. secondary controls		
6-Methyl-5-heptene-2-one	1.36E-06	Primary control
6,6-Dimethyl-2-methylene-bicyclo[3.1.1]heptane	6.28E-06	Primary control
Butan-2-one	0.003586	Secondary control
Nonanal	0.016482	Secondary control
3-methylsulfanylpropanal	0.016482	Primary control
2-methyl-5-(1-methylethyl)-1,3-cyclohexadiene	0.017088	Secondary control
(1R,4E,9S)-4,11,11-trimethyl-8-methylidenebicyclo[7.2.0]undec-4-ene	0.031504	Secondary control
4,7,7-trimethylbicyclo[4.1.0]hept-2-ene	0.041351	Secondary control

*Adjusted for multiple comparisons using false discovery rate (FDR)

Supplementary table 3. Significant correlations identified between fungal operational taxonomic units (OTUs) and volatile organic compounds on integrated analysis. Correlation measured using Pearson's correlation coefficient.

OUT	Volatile organic compound	Correlation	Adjusted P value
FOTU391 - *Saccharomyces mikatae*	Octane	-0.528	0.036
FOTU19 - *Saccharomyces mikatae*	Octane	-0.598	0.005
FOTU240 - *Saccharomyces mikatae*	Octane	-0.596	0.005
FOTU19 - *Saccharomyces mikatae*	Heptane, 2,4-dimethyl-	-0.606	0.005
FOTU240 - *Saccharomyces mikatae*	Heptane, 2,4-dimethyl-	-0.601	0.005
FOTU15 - *Saccharomyces mikatae*	Heptane, 2,4-dimethyl-	-0.542	0.025
FOTU18 - *Saccharomyces mikatae*	Heptane, 2,4-dimethyl-	-0.559	0.016
FOTU391 - *Saccharomyces mikatae*	1,3-Xylene	0.617	0.005
FOTU19 - *Saccharomyces mikatae*	1,3-Xylene	0.653	0.002
FOTU8 - *Candida ethanolica*	3-methylbutanoic acid	-0.611	0.005
FOTU391 - *Saccharomyces mikatae*	1,3-ditert-butylbenzene	-0.592	0.005
FOTU19 - *Saccharomyces mikatae*	1,3-ditert-butylbenzene	-0.756	0.000
FOTU240 - *Saccharomyces mikatae*	1,3-ditert-butylbenzene	-0.541	0.025
FOTU18 - *Saccharomyces mikatae*	1,3-ditert-butylbenzene	-0.601	0.005

Supplementary table 4. Primers used for PCR

Name	Sequence 5’ to 3’	Marker	Reference
Forward overhang	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT	Na	(20)
Reverse overhang	GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T	Na	(20)
FungiQuant-F (forward)	GGR AAA CTC ACC AGG TCCA G	18S rRNA	(21)
FungiQuant-R (reverse)	GSW CTA TCC CCA KCA CGA	18S rRNA	(21)

Supplementary table 5. Read numbers per OTU per sample, summary statistics and OTU taxonomy.

Sample ID	Number of OTUs	Total number of reads	Median reads	IQR
S28-PD24	13	36062	96	618
S62-PD38	13	37433	58	323
S30-IBD-049-18S	22	751903	160.5	911.5
S31-IBD-050-18S	17	169658	68	2171
S66-PD02	8	34777	83	451.75
S76-PD36	10	24438	213.5	4751.5
S52-PD14	11	42077	19	40.5
S75-PD13	9	41324	226	5025
S64-PD17	11	35723	12	293
S36-IBD-057-18S	13	154636	26	145
S41-PD19	13	76697	22	59
S53-PD40	10	82923	25	123.75
S16-PD15	12	33875	33	251.5
S5-PD25	17	42590	48	2407
S17-PD33	14	92724	30.5	70.5
S51-PD04	13	63405	54	617
S65-PD06	9	70704	21	58
S29-PD16	17	48850	15	114
S24-IBD-034-18S	13	228505	52	153
S50-IBD-084-18S	14	199590	195	416
S63-PD09	15	67375	16	250.5
S30-PD26	5	4057	146	1278
S42-PD23	10	42878	22	506.75
S16-IBD-023-18S	17	476737	420	421
S25-IBD-036-18S	16	480218	51	338
S32-IBD-051-18S	12	75326	10	63.5
S48-IBD-082-18S	17	146190	21	311
S52-IBD-087-18S	13	412524	68	349
S54-PD08	12	47708	31	44
S18-PD41	18	44326	15.5	46
S38-IBD-065-18S	10	31714	6.5	44.75
S40-IBD-068-18S	17	288899	30	308
S28-IBD-045-18S	14	108427	22	73.5
S3-IBD-005-18S	13	170318	57	357
S49-IBD-083-18S	11	182722	121	1549.5
S11-IBD-018-18S	10	26404	16	38.5
S33-IBD-054-18S	14	63926	27.5	110.75
S53-IBD-088-18S	11	71035	16	49.5
S74-PD05	14	20520	4	11.25
S38-PD01	9	31088	20	345
S6-PD32	8	13618	130.5	264.5
S40-PD03	13	32181	288	4314
S5-IBD-008-18S	9	50751	7	69

S34-IBD-055-18S	8	36586	34	58.25

See additional supplementary material for full document.